DOCUMENT RESUME

ED 336 425 TM 017 254

AUTHOR Redfield, Doris L.; And Others
TITLE The Role of Student Outcomes in Dual Purpose Teacher Evaluation Systems: A Model for Meeting Top Down and Bottom Up Needs.
PUB DATE Oct 89
NOTE 15p.; Paper prepared for presentation at the Annual Meeting of the American Evaluation Association (San Francisco, CA, October 19-21, 1989).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Accountability; *Educational Assessment; Educational Improvement; Elementary Secondary Education; *Evaluation Methods; Formative Evaluation; *Models; Needs Assessment; *Outcomes of Education; *Pilot Projects; State Programs; Summative Evaluation; *Teacher Evaluation; Testing Programs
IDENTIFIERS Georgia; Kentucky; Teacher Performance Appraisal System

ABSTRACT

A model is presented for including student outcome measures on teacher evaluation systems when there are needs for both formative and summative evaluation data. The experiences of Kentucky and Georgia provided bases for the development of the dual purpose assessment model. Pilot tests in Kentucky and Georgia were specific to the top-down accountability legislation that drove them, but the demands for bottom-up outcome information that could be used for local improvement provided lessons that contributed to model development. The nine-step teacher productivity appraisal process used in Georgia is outlined. The model calls for measures that can defensibly hold teachers accountable to the public and to policy-making groups for particular student achievements. The model also calls for measures that can defensibly hold teachers accountable to themselves, their students, and parents for providing appropriate instruction. The model is based on a broad definition of student achievement to include a variety of cognitive and non-cognitive outcomes. Figure 1 illustrates specific and general outcomes. Figure 2 is a flowchart of the dual-purpose evaluation model. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
DORIS REDFIELD
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 336425

THE ROLE OF STUDENT OUTCOMES IN DUAL PURPOSE TEACHER EVALUATION SYSTEMS: A MODEL FOR MEETING TOP DOWN AND BOTTOM UP NEEDS

Doris L. Redfield

Center for the Study of Evaluation, UCLA Graduate School of Education

James R. Craig

Psychology Department, Western Kentucky University

Jess Elliott

Personnel Development Division, Georgia Department of Education

---

Paper presented at Evaluation '89, the annual meeting of the American Evaluation Association, San Francisco, CA.

October, 1989

TM017254

BEST COPY AVAILABLE

THE ROLE OF STUDENT OUTCOMES IN DUAL PURPOSE TEACHER EVALUATION SYSTEMS: A MODEL FOR MEETING TOP DOWN AND BOTTOM UP NEEDS

Doris L. Redfield, James R. Craig, & Jess Elliott

Public demands for educational accountability at local, state, and national levels are ever increasing. An accompanying assumption seems to be that accountability data can be used to inform policy decisions regarding the improvement of teaching and learning. As public demands for accountability increase, educators are emphasizing their needs for information and resources that can help them provide appropriate instruction to an overwhelmingly diverse population of students. Providing such instruction, they argue, warrants the exercise of professional decision making and autonomous action -- action that seems threatened by various accountability mandates. For example, state-wide testing is viewed by many as having an inappropriate influence on classroom curriculum and instruction.

Often, meeting the assessment needs of accountability and autonomy demands are debated as if they are mutually exclusive enterprises. These debates, whether they occur in public, professional, or political arenas, become particularly heated and complex when they center on the use of student assessment data in the evaluation of teachers.

The purpose of this paper is to present a model for defensibly including student outcome measures in teacher evaluation systems that have simultaneous needs: (a) a need for formative evaluation that can appropriately inform instruction and that requires autonomous decision

making on the part of teachers and (b) a need for summative evaluation that may be used to hold teachers accountable to the public and policy making/governing groups for particular student outcomes. When summative evaluation data are used for accountability purposes, the decisions associated with the data often carry high stakes consequences (e.g., promotion and salary decisions). The experience of two states, Kentucky and Georgia, provided important bases for the development of the dual purpose assessment model described in this paper.

The Kentucky and Georgia Experiences

In 1986-87, Kentucky piloted the first year of a study designed to explore possibilities for including student achievement data in a career ladder plan while avoiding indefensible and inappropriate uses of standardized achievement test scores (Redfield, 1988a). Components of the plan, in addition to student achievement, included teachers' (a) observed instructional performance, (b) professional development activities, and (c) evidence of professional leadership/initiative.

The researchers charged with studying the student achievement aspect of the plan were particularly concerned that measures of student achievement are most often conceptualized as scores on standardized achievement tests. Such test scores, in isolation, cannot be used to defensibly evaluate teachers for various reasons which include the following:

1. Standardized achievement tests are designed to reliably assess students' performance, not teachers' effectiveness.
2. Not all teachers teach subject matter measured by readily available or commonly used standardized achievement tests.
3. There are educational outcomes which are valued by teachers and

parents but which are not measurable using traditional standardized tests (e.g., critical thinking, motivation, self-discipline, self-esteem, positive attitudes, prosocial behaviors).

4. Reasonable expectations of student achivement vary. Average performance or gain is not a defensible expectation for non-average students (e.g., handicapped, disadvantaged, gifted).

5. Many factors influence student achievement that are not under the control of teachers. For example, teachers do not control innate ability or home situations.

6. Unless a teacher is the sole influence on a student's learning, not all of a student's achievement may be attributed to that particular teacher.

Because of the issues involved, Kentucky removed the student achievement component of its career ladder plan for separate study. Data yielded by the separate study on student achievement included (a) designation of the student outcome goals targeted by teachers of different grade levels and subject matter areas; (b) evidence of the extent to which participating teachers were able to document the designated outcomes -- outcomes which included attitudes and behaviors as well as cognitive knowledge and skills; and (c) evidence of the extent to which individual teachers and their supervisors could agree on the priority level of the targeted goals, the difficulty of accomplishing those goals, and the level of goal accomplishment (Craig, Miller, Pankratz, & Redfield, 1988).

In 1987-88, Georgia pilot tested a "teacher productivity" assessment plan that was a logical extension of Kentucky's 1986-87 work (Redfield, 1988b). Teacher productivity was defined by the Georgia program as the "component of the Career Ladder appraisal

process dealing with the academic and behavioral performance of students within the teacher's classroom." This definition was intended to convey the idea that when teachers demonstrate productivity, they are able to provide evidence that their students are making substantial progress -- progress which is related to the academic and behavioral goals and objectives of that particular teacher's class or courses. An important objective of the teacher productivity component of Georgia's teacher appraisal process was to demonstrate that students made substantial progress; the objective was not to estimate the proportion of students' progress that could be attributed to the efforts of any particular teacher. If students made substantial progress, it was assumed that the teacher was a likely, major contributor to that progress.

The Georgia pilot involved (a) developing a training package for use with teachers and supervisors, (b) training teachers and supervisors to formally and systematically implement the documentation procedures previously delineated by participants in the Kentucky study, and (c) developing a teacher productivity scoring rubric to allow for the inclusion of student outcome data in a career ladder accountability system. Outlined below are those aspects of Georgia's teacher productivity pilot that yielded important implications for the dual purpose teacher evaluation model as described in the final section of this paper.

Steps in the Teacher Productivity Appraisal Process

Step 1: Teachers and "supervisors" are trained to use the teacher productivity appraisal process. Supervisors are defined as the persons responsible for evaluating any particular teacher. In most cases, it is the building principal.

Step 2: Each participating teacher drafts a productivity plan. Each plan consists of a set of productivity goals. Based upon data yielded by Kentucky's 1986-87 pilot study, Georgia broadly defined student achievement to include two intersecting categories of achievement outcomes or productivity goals. One category consists of outcomes associated with academic achievements vs. those associated with nonacademic achievements such as behaviors (including attitudes and affects). The other category consists of outcomes that are specific to particular types of students, classes, or courses of study vs. those that are more general in nature and may apply to a wide variety of students, class types, or courses of study. The four categories of achievement outcomes or productivity goals, resulting from the intersection of the two categories, are depicted in Figure 1.

Insert Figure 1 about here

Georgia determined that each teacher's productivity plan should consist of at least three student outcome goals in the category most relevant to his/her teaching assignment and at least one goal in each of the remaining three categories. Each teacher and his/her supervisor jointly determine the most relevant category for that teacher's particular situation. Productivity goals may be targeted at individual students, groups of students, an entire class, or multiple classes.

Step 3: Teachers and supervisors meet to agree upon and finalize the teacher's productivity plan. As part of this activity, the teacher and supervisor use a 5-point scale to negotiate agreement

on the appropriateness and significance of each goal, the role of teacher effort in attaining each goal, and the relationship between each goal and its proposed documentation.

Step 4: Each teacher-supervisor team submits the productivity plan to a review team. The number and kinds of individuals making up the review teams is proposed by each local district and subject to state approval.

Step 5: Based upon the productivity plan submitted by the teacher-supervisor team, the review team decides to approve the plan, disapprove the plan, or approve the plan subject to modifications in areas identified by the review team. In those cases where the teacher and supervisor cannot agree on a plan, the review team serves as an arbitrator. The teacher or supervisor may appeal the review team's decision according to the local district's approved appeals process.

Step 6: Each teacher implements the approved productivity plan. Technical support is to be provided as necessary. To facilatate the documentation and review processes, the paperwork associated with documenting student performance is limited (Redfield, 1988a; 1988b).

Step 7: Near the end of each annual appraisal period, each teacher presents his/her documentation of student outcomes to the supervisor. The teacher-supervisor team finalizes the documentation for presentation to the review team. In preparing documentation for presentation to the review team, the teacher and supervisor use a 5-point scale to reach agreement on the accuracy with which the agreed-upon plan for each goal was implemented, the quality of goal documentation, and the extent to which each goal was attained.

A weighted scoring system is then applied to the points assigned

to each goal by the teacher-supervisor teams. The details of the system are provided elsewhere (Redfield, 1988b); key features of the weighted scoring system are as follows:

o The weighting of the scoring system assumes that goals in the most relevant achievement category are relatively important by virtue of their predetermined relevance. Additionally, or conversely, if a teacher's various productivity goals are not of equal or assumed importance, some goals may be emphasized over others by assigning different weights to them.

o Increasing levels of career ladder status require increased levels of teacher performance. For example, to qualify for Level III status in the teacher productivity aspect of the overall appraisal, teachers must obtain a minimum of 20 points out of 100 possible points across the three-year appraisal period. For advancement to Level IV, 40 points are required; 60 points are required for advancement to Level V. Across years, the points constitute ordinal level data only (i.e., 40 points do not represent twice as much productivity as 20 points, etc.).

o Applicants for Level III, IV, and V career ladder status must be able to document significant productivity (i.e., "substantial" student performance for at least two of the three years of the appraisal period). This provision allows a teacher to have an "off" year due to circumstances beyond his/her control while at the same time demanding an appropriate level of overall productivity.

Step 8: Based on the student achievement documentation submitted by the teacher-supervisor team, the review team decides to (a) recommend a particular career ladder status, (b) call for a clarification conference with the teacher and/or supervisor, or (c)

require additional information. In those cases where the teacher and supervisor have been unable to reach agreement, the review team will consider information provided by both parties. The review team's decision may be appealed by the teacher or supervisor.

Step 9: At the beginning of each year, results of the previous year are used by teacher-supervisor teams to plan for the ensuing year. At the end of the three year appraisal period, the weighted scoring system is used to combine results across years. The combined, weighted score is used in conjunction with scores from the other career ladder components to recommend career ladder status.

Implications

The Kentucky and Georgia pilot tests were specific to the "top-down" accountability legislation that drove them. Nonetheless, the demands of these two pilots for "bottom-up" autonomous action (i.e., the provision of student outcome information that could be used for local improvement at the classroom level) provided some valuable lessons that have generalizable implications. Those implications are reflected by the model described in the next section of this paper.

The Model

It is imperative that accountability models consider WHO is being held accountable, TO WHOM, FOR WHAT (McDonnell, 1989). The components of the following model require consideration of two kinds of accountability; in each case, attention is given to WHO, TO WHOM, and FOR WHAT.

On the one hand, the model calls for measures that can defensibly hold teachers accountable to the public and to policy making/governing groups for particular student achievements. When appropriate, such measures might include standardized achievement test

scores; but, the overall accountability measure should not be limited to these scores. In fact, the use of multiple indicators is cardinal.

The kinds and levels of achievements for which teachers are held accountable and the consequences attached to the accountability data are not technical decisions; they are philosophical and political policy decisions. Of course, such decisions will affect the technical applications of the model in differing situations (e.g., local districts).

On the other hand, the model calls for measures that can defensibly hold teachers accountable to themselves, their students, and the parents of their students for providing appropriate instruction. This is the kind of accountability that requires autonomous teacher action and information other than, or in addition to, test scores. Test scores indicate only that students have relative strengths and weaknesses; they do not pinpoint where or why an individual student's learning in particular areas is relatively strong or weak. The kinds of measures that can provide useful, diagnostic information must be necessarily sensitive to the variety of students likely to be present in a teacher's classroom (e.g., differing abilities, language proficiency levels, and background experiences).

Paradoxically, the information required by teachers for the kind of accountability that allows for instructional autonomy can threaten public or administrative perceptions of their "accountability" or competence. The identification of student weaknesses for purposes of determining instructional remediation are all-too-often interpreted as "low scores" for teachers -- low scores that can result in inappropriate, negative decisions (e.g., do not promote). Figure 2

illustrates a model for fairly and defensibly including student outcome data in teacher evaluation systems when it is desirable that the system meet two needs: (a) top-down needs for public accountability and decision making and (b) bottom-up needs for information that can meaningfully inform classroom instruction.

Insert Figure 2 about here

The model presented in Figure 2 is based upon a broad definition of student achievement which includes a variety of cognitive and noncognitive outcomes. It is also based on a two-fold definition of accountability: (a) teacher accountability to the public for particular student outcomes and (b) teacher accountability to themselves, students, and parents for providing maximally appropriate and effective classroom instruction -- instruction that requires autonomous decision making and action. In conceptualizing this dual-purpose teacher evaluation model, essential considerations have been (a) each stakeholder group's purpose for evaluation, (b) the kinds of information needed by each group for appropriate decision making, (c) data gathering procedures that can be shared across groups versus those that cannot, and (d) the unique needs of each group in receiving meaningful information in usable form. Provisions for training and technical support are integral aspects of the model.

References

Craig, J.R., Miller, S.K., Pankratz, R., & Redfield, D.L. (1988). Kentucky Career Ladder Commission research report on the 1986-87 pilot program. ERIC Reproduction Service, ED 299240.

McDonnell, L. (1989, October). State and district issues: The role of indicators and assessment in school reform and school restructuring. Presentation, UCLA/CRESST Conference entitled "Educational Quality Indicators: Taking Stock," Los Angeles, CA.

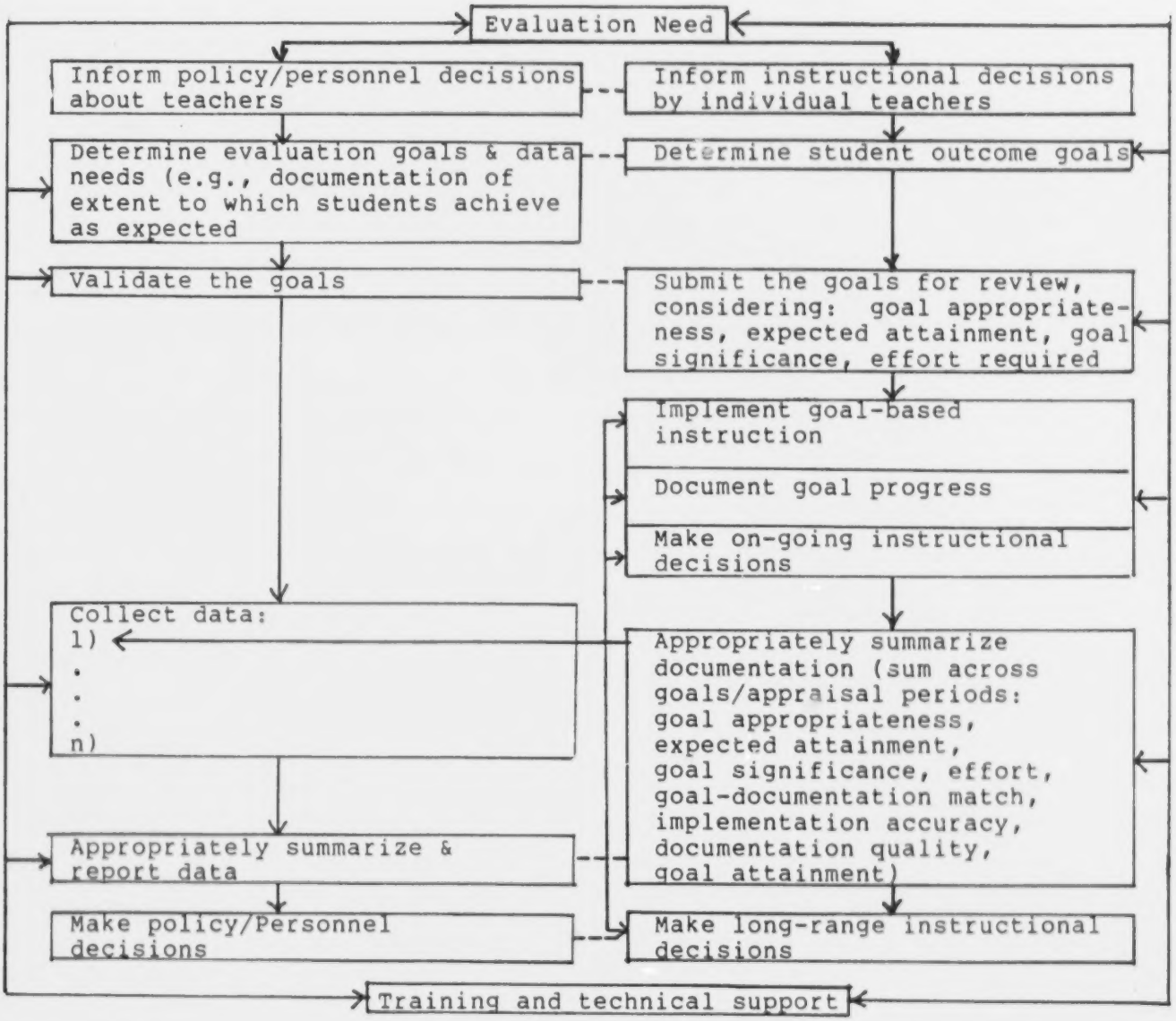
Redfield, D.L. (1988a, April). Expected student achievement and the evaluation of teaching. Paper presented in G. Galluzzo (Chair), Instrumentation for evaluating teachers for professional advancement: Development and research. Annual meetings of the American Educational Research Association, New Orleans, LA.

Redfield, D.L. (1988b). Guide: Teacher productivity appraisal process. Available from D.L. Redfield, UCLA Graduate School of Education, Los Angeles, CA. or J. Elliott, Georgia Department of Education, Atlanta, GA.

Figure 1

Scope of Outcome	Type of Outcome: Academic	Type of Outcome: Nonacademic
Specific (to type of class or course of Study)	Category: Academic Specific	Category: Nonacademic Specific
General (applies to a wide variety of class types and courses of study)	Category: Academic General	Category: Nonacademic General

Figure 2


Evaluation Need
Inform policy/personnel decisions about teachers
Inform instructional decisions by individual teachers
Determine evaluation goals & data needs (e.g., documentation of extent to which students achieve as expected
Determine student outcome goals
Validate the goals
Submit the goals for review, considering: goal appropriateness, expected attainment, goal significance, effort required
Implement goal-based instruction
Document goal progress
Make on-going instructional decisions
Collect data:
1)
.
.
.
n)
Appropriately summarize documentation (sum across goals/appraisal periods: goal appropriateness, expected attainment, goal significance, effort, goal-documentation match, implementation accuracy, documentation quality, goal attainment)
Appropriately summarize & report data
Make policy/Personnel decisions
Make long-range instructional decisions
Training and technical support

Note: ----- = separate but parallel activities
_____ = shared activities